SCIENCE CHINA
Life Sciences

• SHORT COMMUNICATIONS •

January 2011 Vol.54 No.1: 65–67
doi: 10.1007/s11427-010-4114-1

Tetrahymena Gene Expression Database (TGED): A resource of microarray data and co-expression analyses for *Tetrahymena*

XIONG Jie[1,4†], LU XingYi[1,4†], LU YuMing[1†], ZENG HongHui[1†], YUAN DongXia[1], FENG LiFang[1,4], CHANG Yue[1,4], Josephine BOWEN[3], Martin GOROVSKY[3], FU ChengJie[2] & MIAO Wei[1,2*]

[1]*Key Laboratory of Aquatic Biodiversity and Conservation, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China;*
[2]*State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China;*
[3]*Department of Biology, University of Rochester, Rochester, NY 14627, USA;*
[4]*Graduate University of Chinese Academy of Sciences, Beijing 100049, China*

Received July 5, 2010; accepted October 12, 2010

Tetrahymena thermophila is a model eukaryotic organism. Functional genomic analyses in *Tetrahymena* present rich opportunities to address fundamental questions of cell and molecular biology. The *Tetrahymena* Gene Expression Database (TGED; available at http://tged.ihb.ac.cn) is the first expression database of a ciliated protozoan. It covers three major physiological and developmental states: growth, starvation, and conjugation, and can be accessed through a user-friendly web interface. The gene expression profiles and candidate co-expressed genes for each gene can be retrieved using Gene ID or Gene description searches. Descriptions of standardized methods of sample preparation and the opportunity to add new *Tetrahymena* microarray data will be of great interest to the *Tetrahymena* research community. TGED is intended to be a resource for all members of the scientific research community who are interested in *Tetrahymena* and other ciliates.

***Tetrahymena thermophila*, gene expression database, co-expression**

Citation: Xiong J, Lu X Y, Lu Y M, *et al*. *Tetrahymena* Gene Expression Database (TGED): A resource of microarray data and co-expression analyses for *Tetrahymena*. Sci China Life Sci, 2011, 54: 65−67, doi: 10.1007/s11427-010-4114-1

The ciliated protozoan *Tetrahymena* has been used as a model eukaryotic cell since Nobel Laureate Andre Lwoff first placed it into pure culture in 1923. This organism has contributed to the understanding of basic eukaryotic mechanisms, including the 1989 Nobel Prize-winning discovery of catalytic RNA [1], and the 2009 Nobel Prize-winning description of telomere structure and telomerase [2]. The most intensely studied species of *Tetrahymena*, *T. thermophila*, is easily grown and maintained using standard microbiological techniques, without specialized apparatus [3]. Its life cycle permits the use of molecular genetic tools, such as gene knockout and gene over-expression [4].

T. thermophila belongs to the phylum Ciliophora in taxonomy, and this phylum also includes the generus known to us which is called *Paramecium.* Like many other ciliates, *T. thermophila* exhibits the nuclear dimorphism. Each *T. thermophila* cell has two nuclei, the germline micronuclear (MIC) and the somatic macronuclear (MAC) [3]. The MICs, containing five chromosomes, fuse to produce a new zygotic MIC in conjugation of two mating types, and the mitotic copies of zygotic MIC will produce a new MAC in

†Contributed equally to this work
*Corresponding author (email: miaowei@ihb.ac.cn)

conjugation [5]. Although the MAC is derived from the MIC, it is distinct from it, and has been estimated to contain about 225 chromosomes. The MAC is transcriptionally active throughout the life cycle of *T. thermophila* [6]. The completion of the *T. thermophila* Macronuclear Genome Project [6], establishment of the *T. thermophila* Genome Database (TGD; www.ciliate.org) [7], development of molecular genetic technologies for DNA-mediated transformation by homologous integration and a genome-wide microarray platform [8] have enabled functional genomic analyses in *Tetrahymena* and allow fundamental questions of cell and molecular biology to be addressed.

The *Tetrahymena* Gene Expression Database (TGED) was established based on microarray data of gene expression in *T. thermophila* cells during growth, starvation, and conjugation. In addition to providing expression profiles and candidate co-expressed genes of most genes in *T. thermophila*, TGED introduces standardized methods of sample preparation and provides a platform for the community to share *Tetrahymena* microarray data.

1 Contents of the TGED

The TGED integrates expression data from 50 individual genome-wide arrays [8]. These data cover three major physiological and developmental states in *Tetrahymena*, including three stages of growth, seven stages of starvation, and 10 stages of conjugation.

Using the Roche NimbleGen Systems, the 2006 annotation version of the *T. thermophila* macronuclear genome in the TGD was used to design the genome-wide microarray. Arrays were scanned and normalized using NimbleScan software. Normalized expression values for the individual probes were used to obtain the expression values for a given open reading frame (ORF), using the robust multiarray average (RMA) procedure [9]. Finally, the data were analyzed based on the RMA-processed expression values (RMA calls). The RMA calls of *T. thermophila* during growth, starvation, and conjugation were obtained via the Gene Expression Omnibus, and then were integrated and exported as a data file using ArrayStar software, version 2.0 (DNASTAR, Inc., Madison, WI, USA). This file was used for the TGED, and 20 data points were used for the graphic expression profile of each gene.

Genes involved in the same biological process often display coordinate expression profiles; therefore, co-expression data can provide valuable information for analyzing regulatory relationships and indicating protein-protein interactions (PPI) [10]. Using all the data from growing, starved, and conjugating cells, pair-wise Pearson correlation coefficients (R) were calculated for *T. thermophila* genes with a program compiled by C++ (see Appendix in the electronic version). Correlation coefficients with a value of more than 0.8 are indicative of co-expression. The co-expression analyses were shown to identify proteins in known protein complexes that function in the same developmental process (DNA elimination) during conjugation [8].

2 The TGED web interface

The TGED can be accessed through a user-friendly web interface. There are two main ways of searching the database: Gene ID or by gene description.

For Gene ID searches, the SEARCH genes interface allows the user to input a Gene ID in the 2006 version of the TGD format (ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/t_thermophila/annotation_dbs/interim_annotation_release_08313006/), which returns detailed information on the gene, its expression patterns, and co-expressed genes (Figure 1). The information returned for each gene includes the Gene ID (hyperlinked to the NCBI), the Sequence ID (unique ID in the microarray design), the Gene Type (protein-coding, non-protein-coding, or tRNA), the RNA name (gene description), and Sequence (full length cDNA, protein, and the 13–14 probes used in the microarray design) (Figure 1B). The Expression Profile shows a graph with the expression value (arbitrary units, AU) from 20 stages of growth, starvation, and conjugation (Figure 1C). The Co-expressed Genes lists all candidate genes with expression patterns similar to the selected gene and the correlation coefficients between them (Figure 1D). The user can select any one of the genes and find its expression pattern and co-expressed genes.

Keyword searches are also allowed, in which case the input is matched against the gene name and description in the 2006 version of the TGD gene annotation. The user may choose from among a number of returned candidates. As before, the information, expression pattern, and co-expressed genes of the selected gene will be provided.

3 Other services in the TGED

The page “Sample Preparation” describes the standardization of culture conditions for growth, starvation, and conjugation, the method for preparing RNA for the microarray, and the standard Bioanalyzer profile of *Tetrahymena* RNA for quality control. The pages “Submit” and “Download” provide space for colleagues to communicate and share other *Tetrahymena* microarray data. The user can submit the “title”, “summary”, “overall design”, and “contributor” of his own *Tetrahymena* microarray data in the Submit interface, and download other microarray data after getting the ID and password from the TGED administrator in the Download interface.

4 Future developments

Future developments of TGED will include uploading and

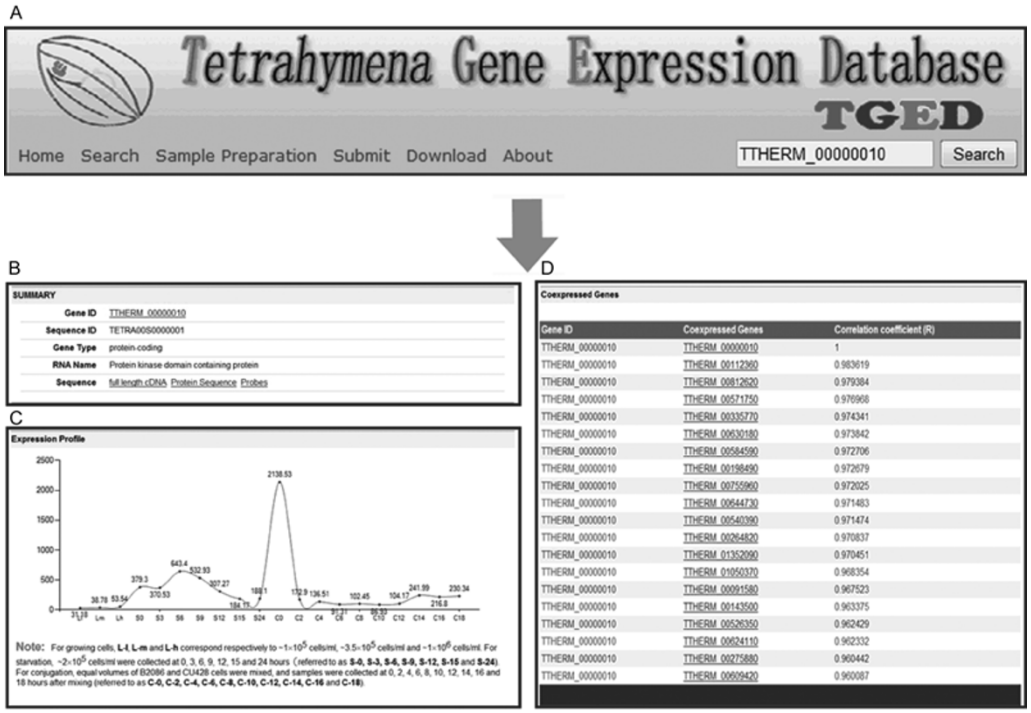

Figure 1 Screenshots of TGED pages relevant to the search for gene TTHERM_00000010. A, Top page. B–D, Search results: summary of gene information (B), expression profile (C), and co-expressed genes (D).

integrating additional *Tetrahymena* expression datasets as they become available. For example, those of *T. thermophila* under stress from exposure to pollutants or displaying the effects of specific gene mutations. In addition, there will be further optimization of algorithms for co-expression analysis, and construction of metabolic pathways and protein interaction and gene regulatory networks in *Tetrahymena*.

This work was supported by the National Natural Science Foundation of China (*Grant Nos. 30870356 and 30970424*), *the Knowledge Innovation Program of Chinese Academy of Sciences* (*Grant No. KSCX2-YW-G-072*), *and the National Institutes of Health* (*Grant Nos. GM021793 and GM072752*).

1 Kruger K, Grabowski P J, Zaug A J, *et al*. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell, 1982, 31: 147–157
2 Greider C W, Blackburn E H. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell, 1985, 43: 405–413
3 Asai D J, Forney J D, Willson L, *et al.* eds. *Tetrahymena thermophila*. Methods in Cell Biology, Vol. 62. New York: Academic press, 2000
4 Turkewitz A P, Orias E, Kapler G. Functional genomics: The coming of age for *Tetrahymena thermophila*. Trends Genet, 2002, 18: 35–40
5 Martindale D W, Allis C D, Bruns P J. Conjugation in *Tetrahymena thermophila*. A temporal analysis of cytological stages. Exp Cell Res, 1982, 140: 227–236
6 Eisen J A, Coyne R S, Wu M, *et al*. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. PLoS Biol, 2006, 4: e286
7 Stover N A, Krieger C J, Binkley G, *et al*. Tetrahymena Genome Database (TGD): A new genomic resource for *Tetrahymena thermophila* research. Nucleic Acids Res, 2006, 34: D500–503
8 Miao W, Xiong J, Bowen J, *et al*. Microarray analyses of gene expression during the *Tetrahymena thermophila* life cycle. PLoS One, 2009, 4: e4429
9 Irizarry R A, Hobbs B, Collin F, *et al*. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics, 2003, 4: 249–264
10 Bhardwaj N, Lu H. Correlation between gene expression profiles and protein-protein interactions within and across genomes. Bioinformatics, 2005, 21: 2730–2738